Pre-Hire Task

Professor Jordan J. Norris*

2022

1 Preamble

- Key attributes I want see during the task
  - Keep in mind throughout that this role of research assistant exists because I do not have time to complete the tasks myself. Efficient outsourcing of these tasks to you is the goal. Anything you can do to minimize time burden on me is highly valued. The following attributes are some of those that are important in achieving this.
  - You are able to complete tasks accurately and you follow instructions carefully. Not reading the task properly or missing some elements of the tasks in your submission is not acceptable.
  - I want to see that you are able to organise your output (code, exhibits, reports, data, etc) well. They need to be structured, neat, clean, readable, etc (the template .do file in the code directory is an example of this for coding in Task 1).
    - Figures must have titles, axes labels, etc. Data variables should be labeled. A degree of beauty in your work is desirable.
    - Be OCD about all this. When we need to revisit earlier code/output months from now, we will thank ourselves for this. And more so if you are no longer on the project and your successor needs to pick up where you left off. But obviously don't waste excessive time on this either.
    - See https://web.stanford.edu/~gentzkow/research/CodeAndData.xhtml for additional guidance on this (not all is relevant but ultimately you'd be expected to familiarize yourself with it when working with me).

*New York University Abu Dhabi, jjnorris@nyu.edu

  - Take initiative. The following is what will be expected when working for me.
    * Be involved in the research process and be mindful about what we are doing in the project.
    * Sometimes the instructions I give might be vague or not work exactly as I've suggested. Slight amendments can be required and you are encouraged to try these things. It's a balance: veering too far from the instructions without speaking to me may be a waste time.
    * Of course I am here for guidance but if I have to hold your hand on every step, then it will be more efficient for me just to complete the task myself. And the purpose of the RA position will have failed.

- Dropbox directory
  - Directory structure[1]
    * Code: scripts (e.g. .do, .m files)
    * Raw: only raw data, do not save any modified files here
    * Temp: save intermediate output
    * Output: save final output (log files, graphs, final datasets)
    * Log: log files
    * Report: save reports in here
    * Task: instructions
  - Relative paths: in all your code, you must use relative paths (see the explanation in task 1)

2 Task 1: Structural Change

- You will investigate *structural change* at the city level in the UK
  - Countries develop over time: agriculture → manufacturing → services
  - See figure 1 for empirical evidence on this
  - As countries get richer (to the right on x-axis), their agricultural employment share declines, the service share rises, and the manufacturing share peaks in the middle

[1] This structure is taken from https://web.stanford.edu/~gentzkow/research/CodeAndData.xhtml.

- 1) Create dataset
  - employment_city_sector.csv: sectoral employment share of total employment
    - 3 sectors: agriculture, manufacturing, services
    - Define "services" to be: retail sector, the financial sector, business administration, leisure and cultural activities
      - (UK parliament website stating the composition of the service sector: https://commonslibrary.parliament.uk/research-briefings/sn02786/)
  - gdp_pop_city.xlsx: gdp, population
- 2) Regression analysis
  - Regress by OLS sectoral employment share on log gdp per capita
  - Do for each of the three sectors
  - Use robust standard errors
- 3) Create graphs like the employment one in figure 1
  - y-variable: sectoral employment share of total employment
  - x-variable: log gdp per capita
  - Each point will be a city rather than a country, in a single year
  - Add line of best fit with the coefficient on sectoral employment share (this is found in part 2)
    - Write coefficient as *coefficient-magnitude (coefficient standard error)*, round each to 3 decimal places
    - This code will be helpful: *local coeff : di %3.2f _*
  - Label axes (not with variable name e.g. loggdp but Log GDP), add graph title (sector name)
  - x-axes ticks should have e.g. 10^2 not 2.
  - Remove background default blue color: *graphregion(color(white)) bgcolor(white)*
  - *i.e Make the graph look nice and readable.*
  - export as pdf
- Do everything in Stata
  - All data extraction, manipulation, and graphing

  - No manual (by hand) adjustment of the data allowed
  - I've created a skeleton .do file for you complete in code: "analysis.do"
    * The file uses relative paths. e.g. I uncomment my parentDirectory, comment yours, and click do, and everything runs. No other adjustment needed.
    * Comment the code well! Stick to my commenting format in the skeleton.
  - Tips:
    * Google things you don't know how to do e.g. "import excel data in Stata". The resources on there are really good.
    * Once you have the command, type "help [command]" in the Stata command line to get see the usage, syntax, examples, etc (this will also be available on Google).
    * To get you started, the appropriate command for the above example is "import excel".

- log the complete execution of your code
  - begin log: *log using log/task1.log, replace*
  - end log: *log close*

3 Task 2: Matrix Properties

Proposition 1. *For an $N \times N$ matrix, M_{ij}, that is row diagonally dominant*

$$\forall i: \quad M_{ii} > \sum_{j \neq i} M_{ij} \tag{1}$$

and row stochastic

$$\forall i: \quad \sum_{j} M_{ij} = 1 \tag{2}$$

then, the diagonal elements of the inverse of the matrix are all greater than or equal to 1

$$\forall i: \quad \left\{M^{-1}\right\}_{ii} \geq 1 \tag{3}$$

- You will investigate the proposition 1 using simulations (preferably in Matlab, but you can use different software if you feel strongly about it). Specifically, you will confirm to a reasonably degree that it is true.

  - We will do this for $N \in \{2, 5\}$ only.

- Hint on how to complete:
  - You will need to create an algorithm that randomly simulates a matrix that satisfies properties (1) and (2). Set the number of simulations to be 10000.
  - You will then check in each simulation that the matrix also satisfies (3).
  - If you do not find any examples that violate (3), we can reasonably conclude that proposition 1 holds.

- Output of the task:
  - .m file (or equivalent) with all code.
  - In the report folder, create a single pdf containing: for each N, a histogram of the distribution of $\min_i \{M^{-1}\}_{ii}$ across all simulations.
    * The histogram must be well labeled (axes and title).
    * Note that $\min_i \{M^{-1}\}_{ii} > 1$ implies (3). Hence if the histogram falls to the right of 1 on the x-axis, the proposition is not rejected.
  - Log the execution of your code

- Keep in mind: I've not gone into as much detail with layout etc in this task, or given a template. But you still must continue to create your code, output etc within the guidelines I've explained in the Preamble, and as exemplified by task 1.

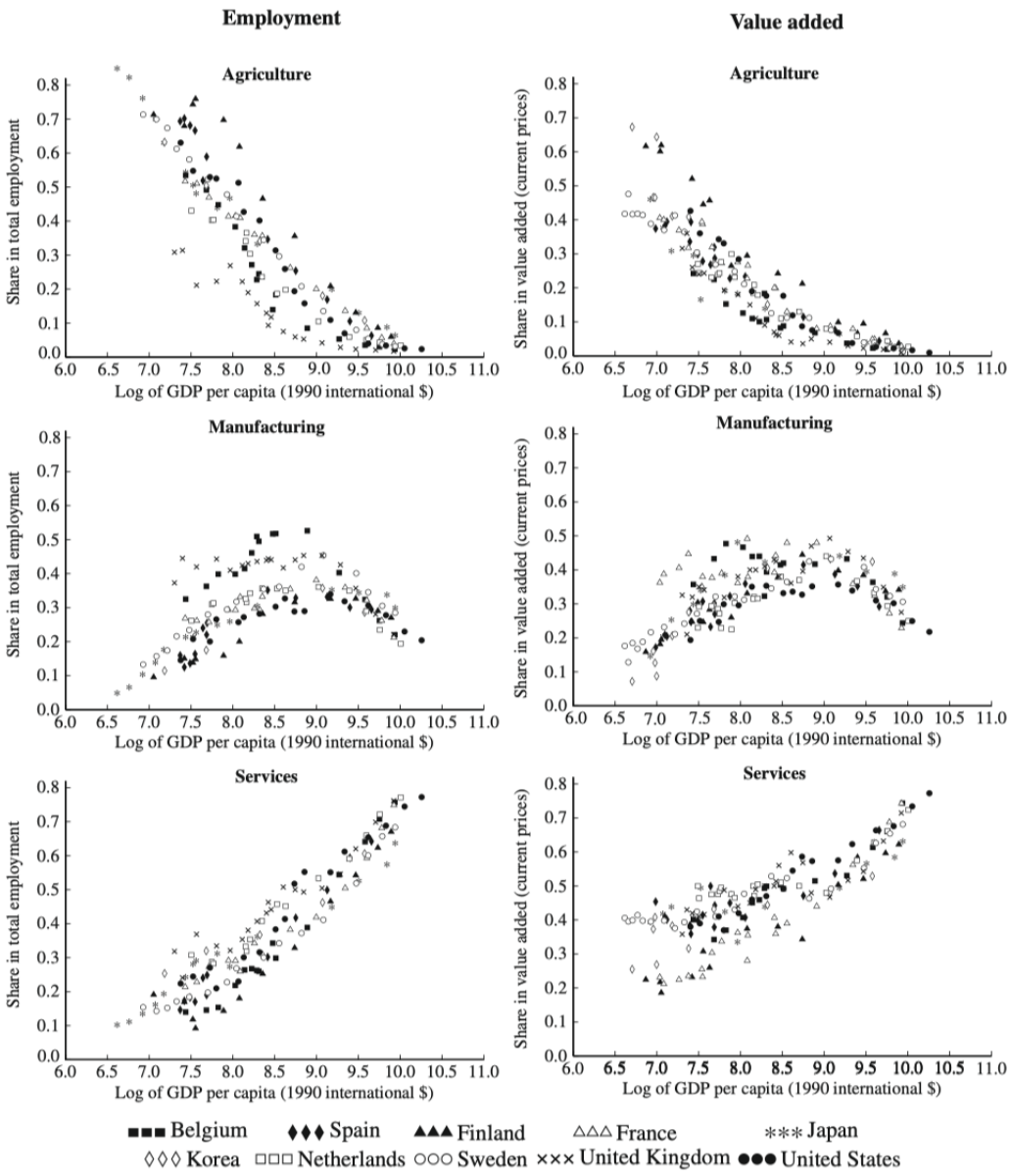

Figure 6.1 Sectoral shares of employment and value added—selected developed countries 1800–2000. *Source: Various historical statistics, see Appendix A.*

Figure 1: Task 1 output to replicate